■   103

Saron Music Transcription Based on Rhythmic Information Using HMM on Gamelan Orchestra

Yoyon K. Suprapto[1], Yosefine Triwidyastuti[2]
[1] Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2] Sekolah Tinggi Manajemen Informatika dan Teknik Komputer (STIKOM), Surabaya, Indonesia
email: yoyonsuprapto@ee.its.ac.id[1], yosefine_its@yahoo.com[2]

Abstract

Nowadays, eastern music exploration is needed to raise his popularity that has been abandoned by the people, especially the younger generation. Onset detection in Gamelan music signals are needed to help beginners follow the beats and the notation. We propose a Hidden Markov Model (HMM) method for detecting the onset of each event in the saron sound. F-measure of average the onset detection was analyzed to generate notations. The experiment demonstrates 97.83% F-measure of music transcription.

Keywords: gamelan music, music tempo, onset detection, hidden markov model, music transcription

1.   Introduction

Gamelan is a traditional musical instrument of Indonesia, which comes from Java. In order to preserve gamelan as national heritage and to bring back the greatness of this music as it was in 17-18 century, some efforts must be conducted to make people more familiar with gamelan and to help them play this instrument easier.

Gamelan consists of about fifteen groups of different instruments, such as Saron, Kenong, Kempul, Kendang, Bonang, etc. Some of the instruments have the same fundamental frequency, such as saron and bonang [1]. See Figure 1. In gamelan music, saron and bonang are not sounded at the same time. Bonang is struck a half beat before saron time, where beat, in this case, is defined as the distance between two consecutive sounds Saron, [2]-[3] See Figure 2.
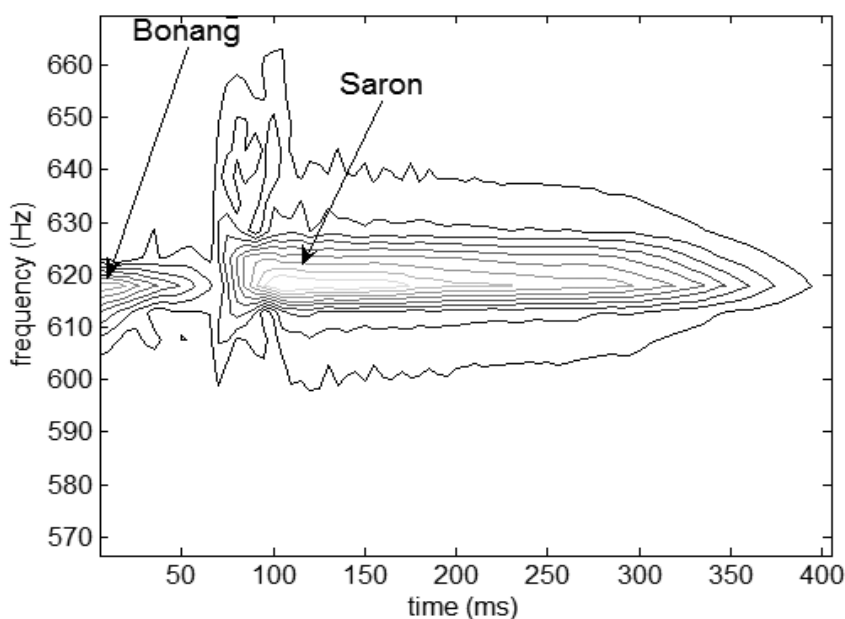

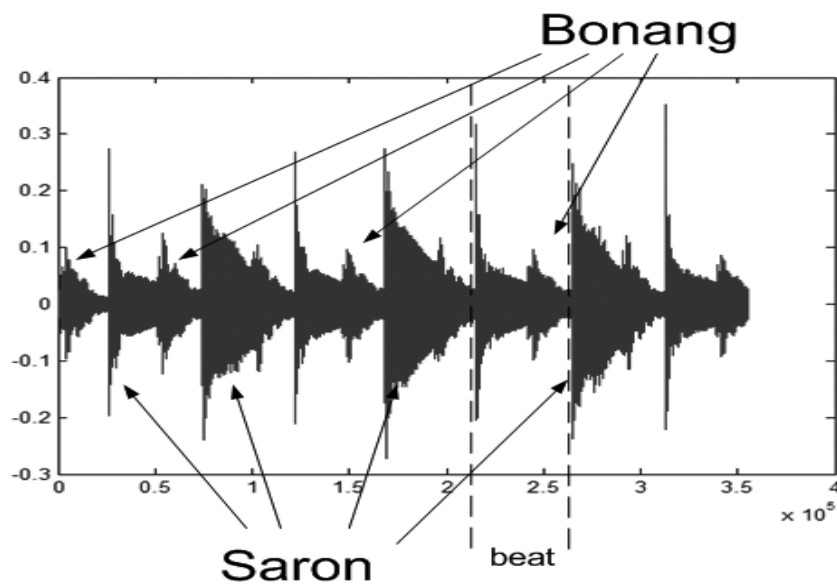Figure 1. Spectrums of bonang and saron

Figure 2. How Saron and Bonang were played

One commonly used method to detect onset is the feature-based onset detection [4]-[8]. The disadvantage of this conventional method is susceptible to weak onset feature or spurious peak that not corresponds to an onset event. See Figure 3.
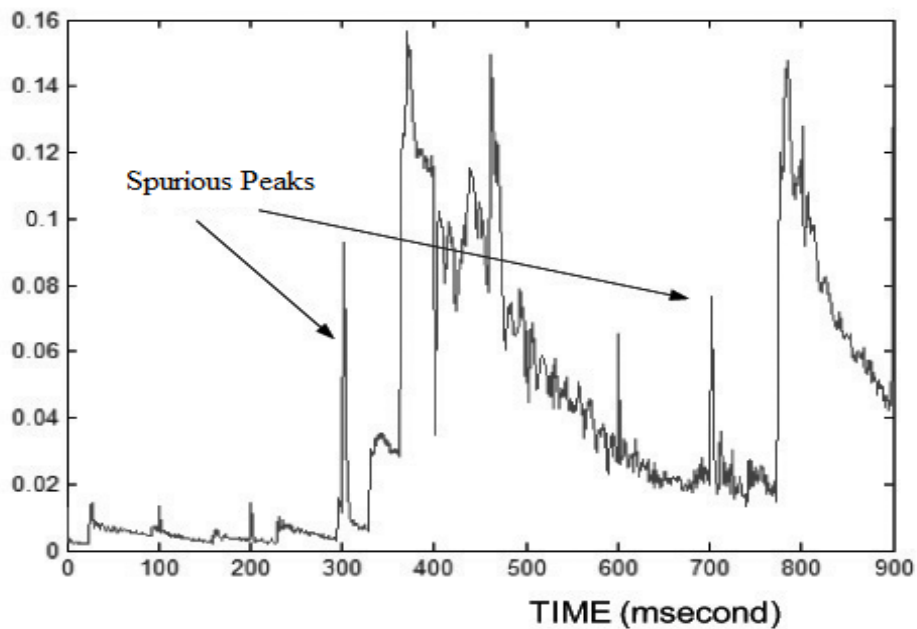

Figure 3. Spurious peaks in Saron magnitude extraction result from a real recorded Gamelan orchestra signal

Another difficulty, gamelan instruments are handmade and were tuned based on the sense of craftsmen. Thus, gamelan music signals often have fluctuations in amplitude, frequency and phase [1]. These fluctuations may lead to different shape of signal envelope.

We propose a Hidden Markov Model (HMM) approach to predict the likely timing of the onsets of gamelan music signals. HMM method allows information such as tempo to be combined in onset detection method [9] - [10]. Tempo information gives prediction for early detection of onsets and discriminates them from false peaks coresponding to another instrument's sound. HMM offers efficient computation and does not need large data for training.

Transcription is essentially done by detecting the onset of a specific type of instrument. Many studies have been carried out to detect the onset of musical events, but those are especially focused on western music. Common onset detection should be improved to detect the onset of the eastern musical instruments such as the gamelan.

Standard onset detection methods are used first to find the location of the peaks by measuring the abrupt change in energy content, magnitude, frequency, or phase of the music signal, then to apply a threshold to decide whether it is a peak or not, by considering height as an onset [5]-[8]. If a peak's height is above the threshold, then the peak will be considered as an onset event and vice versa. This onset detection accuracy only depends on single peak that is analyzed at current time. Therefore, this method could not distinguish the spurious peaks in Bonang signal with the real onset event and the weak onset peak due to many fluctuations in Gamelan music signal such as in Figure 4.

In gamelan ensemble, an instrument sound is always interfered by those of other instruments. For example, the extracted saron sound may still contain bonang sound since both instruments have the same fundamental frequency. But the presence of bonang sound can be distinguished from saron sound by comparing the spectral envelope of both sounds, since bonang sound (60 ms) has shorter envelope than that of saron (300 ms).

In the conventional methods of onset detection, each peak is only evaluated individually without considering the temporal relationship with other peaks. If an onset has appeared, the next onset will not appear in the near future, unless a certain time interval has passed. A clear example is the interval between beats in the form of music that can often be followed by the audience. It's kind of important information that is difficult to relate the onset of feature-based detection.

Other recent onset detection methods are using machine learning. An artificial neural network can be trained to detect the onset of the event [11]-[12]. Important prerequisite for machine learning methods is that the training data must be large enough to represent the actual data in reality. In some cases, the amount of training data should reach up to 70% of all actual data [2]. Due to fluctuations in the signal can be found in many gamelan music, all the variations of the signals must also be included in the training process. In order to detect the onset of the event many different gamelan instruments, the network can not be trained by just one variety of gamelan instruments. Thus, one requires a large database to train the network.

Three stages in onset detection methods can be seen in Figure 3.
1. *Preprocessing,* an optional initial process to accentuate or attenuate some aspects of the original signal is related to the onset detection. At this stage, the signal's spectrogram is divided into multiple frequency bands.
2. *Reduction,* is the most important part in the onset detection, because the original signal is converted into a sampled function of detected onset. In general, the original signal is transformed into the detection function by standard features such as an explicit signal amplitude, frequency or phase
3. *Peak-picking,* which is the final process after the detection function, is formed and onset peaks began to appear. At this stage smoothing or normalization can be done in advance so as to facilitate peak-picking process that identifies the locations of local maxima that are above the threshold.
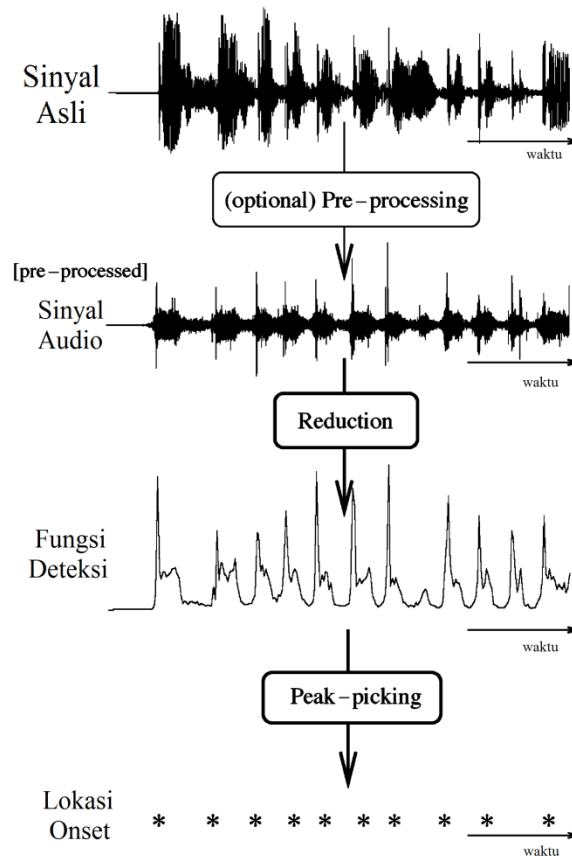
Figure 4. Flowchart Onset Detection Standards [8]

1.1. Previous Detection Onset methods
There are several onset detection methods:

a. Spectral Flux
Method of *Spectral Flux* (SF) is based on the detection of sudden changes in the signal positive energy that shows a part of a new event. *Spectral flux* measures the change in the amount of energy in each frequency bin, and summed giving onset detection function. The formula is written in eq. (1) and (2) [7].

$$FS_{L1}(n) = \sum_k H(|X(n,k)| - |X(n-1,k)|) \text{ (1)} \tag{1}$$

$$FS_{L2}(n) = \sum_k \{H(|X(n,k)| - |X(n-1,k)|)\}^2 \tag{2}$$

with $H(x) = \frac{x+|x|}{2}$ is the half-wave rectifier function and *X(n, k)* is the result of the STFT of the input signal *x* at every n^{th} frame in the kth frequency bin. Based on empirical experiments [7], it is known that the function of L1-norm in eq. (1) is superior to L2-norm [7] in eq. (2).

Selection of Spectral Flux method as a method of comparison due to the characteristics of the equipment Gamelan is a musical instrument played by percussive or beaten. The percussive features of the gamelan instrument, cause changes of magnitude more prominent than the other features [7].

$$f(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|) \tag{3}$$

In this method, we are going to look for the large changes in the output magnitude STFT *X(n, k)* Magnitude of the peak detection method of Spectral Flux is based only on the magnitude of the change in magnitude with the previous frame. In eq. (3), the difference in the magnitude of each frame *n* is rectified, then the result of the kth frequency bin summed up all the window length *N*.

As the process continued, it takes peak-picking process to analyze the magnitude of change in multiple frames at once. Threshold value for detection function at time *t* is the average of the detection function in the analysis window centered at *t*.

$$d(n) = f(n) - \frac{\sum_{k=t-mw}^{t+w} f(k)}{mw+w+1} \tag{4}$$

Then a peak at the n-th frame is selected for the beat or beats if it is a local maximum.

$$d(n) \geq d(k), \quad \forall k: \quad t - w \leq k \leq t + w \tag{5}$$

The selection of the value of *w* is based on the average period of music that indicates the average distance between the knock on the Gamelan music signal. While the variable m is a multiplier variable given a value of 1 so that the determination of a peak based solely on the magnitude of the peak height of the frame to the other frames in a range of music tempo on average.

b. Phase deviation
One of Phase components observed in the detection method is the instantaneous frequency changes that are indicators of possible onsets. Let *φ(n,k)* is the phase component of *X(n,k)* STFT results of the input signal to the nth frame in the frequency bin k.

$$X(n, k) = |X(n, k)| e^{j\varphi(n,k)} \tag{6}$$

φ (n, k) has a range of values from-π to π. So the instantaneous frequency, φ '(n, k), is calculated from the first time-difference of phase spectrum φ'(*n,k*).

$$\varphi'(n, k) = \varphi(n, k) - \varphi(n - 1, k) \tag{7}$$

Then change of the instantaneous frequency can be derived from the second order time-difference of phase spectrum:

$$\varphi''(n, k) = \varphi'(n, k) - \varphi'(n - 1, k) \tag{8}$$

As a final step, the onset detection function based on phase deviation obtained from the absolute value of the average instantaneous frequency changes in all bin frequencies [7].

1.2. Onset Detection Based on spectral Features
In feature-based onset detection, the input signal is converted into the detection function through reduction process by observing the sudden change of the standard features such as audio signal's explicit information on energy (or amplitude), frequency content, or phase. In the following subsections we briefly reviewed the existing approach of onset detection using spectral flux and phase deviation [13]

The rest of the paper is organized as follows. Section II describes proposed method, the HMM method that is used for our onset detection describes the methods used, while Section III describes our whole method of performance measurement, presents our experimental results and discuss the results. The last section concludes this work.

2. Proposed Method
And seThe proposed method in this research is using *Hidden Markov Model (HMM)* for the Saron's beat tracking. In this study, to analyze the performance of the Saron's beat tracking

using HMM methods, we compare it with conventional detectors Spectral Flux, applying the principles of adaptive thresholding. Flowchart of the methods used in this study can be seen in Figure 5.
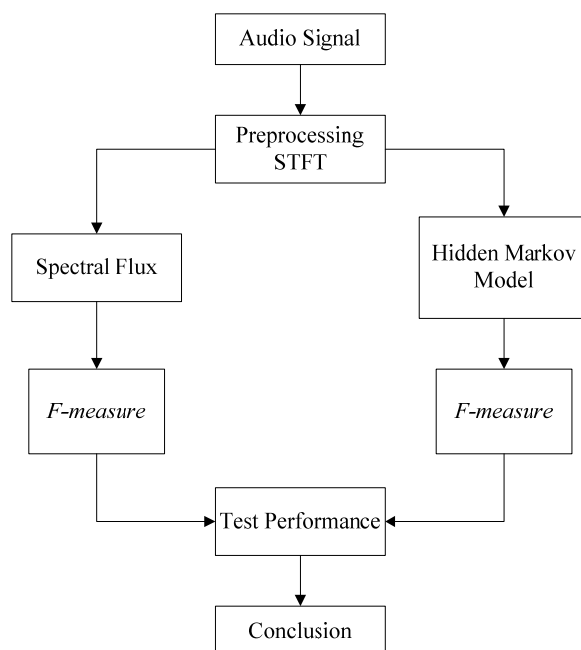

Figure 5. Flowchart of Research Methodology

2.1. Frequency Filtering

In Gamelan orchestra, tactus level corresponds to the speed of beats that sounded by the Saron instrument, which only a single strike to each notation. Therefore, in a system of assessment Gamelan music tempo, frequency filtering processes necessary saron instrument, 500-1000 Hz, especially if the audio signal is a signal observed Gamelan orchestra.

Kaiser window is one way to form a filter. Kaiser filter type can create a wide area and is restricted to a narrow band region.

2.2. Preprocessing

Reviewed Modeling systems is an audio signal that has been processed with overlap Short Time Fourier Transform (STFT). In the STFT, the audio signal is represented in two domains, the time domain and frequency domain.

STFT process can be used to emphasize the feature magnitude of an audio signal. Moreover, the representation in the frequency domain allows the screening process when onset detection of saron is applied to the signal Gamelan orchestra.

This study used a window length $N = 2048$, or equivalent to 43 ms at 48 kHz sampling frequency. Both window length and hope length in overlapped STFT were used to maintain frequency resolution and time resolution respectively. In order to get a smaller index again, the width of hops h = 10 ms (or 76% overlap) is used. The usage of wide-hop size is already commonly used in these studies for detection of the onset and beat tracking [13].

2.3. Hidden Markov Models

The flow chart of Hidden Markov Model (HMM), proposed in this research can be seen in Figure 6. Generally, the method is the incorporation of HMM probability value of observation, transition probabilities, and the initial probability to predict the value of a state that has an order or regular enough structure so that many variations of the observational data can be

approximated by the information structure of the transition state. In this study HMM is used to extract tempo infofmation of musical pieces that will be useful for eliminating the false peaks in transcription.
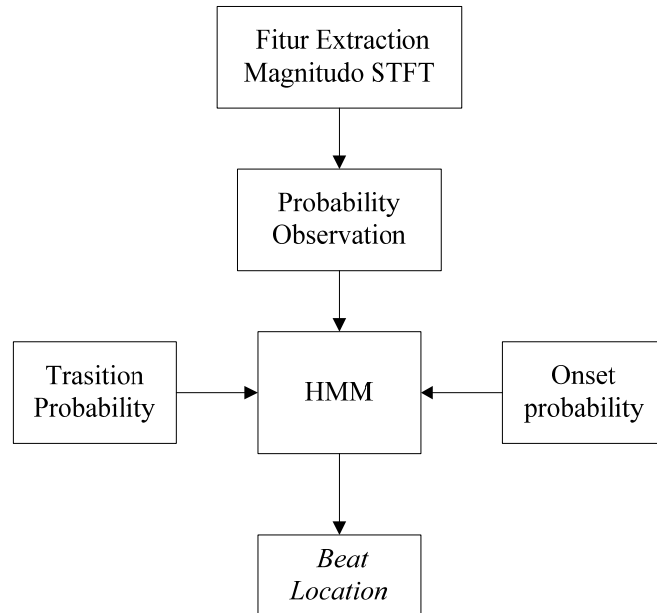

Figure 6. Flowchart of the HMM

Hidden variable τ_t is defined as the number of frames since last onset, which is worth one if the frame is an onset frame (state event = 1). Conversely, if $\tau_t = s$, means that the t - frame is an sth frame from the onset to the last frame. If the detection of the onset of the next frame is met, then the state event back to being equal to 1 and moves up again to encounter the onset of the next frame. Total number of state S that may arise is calculated from the maximum frame spacing between two successive beats.

In each frame, the system also issues the observational data ot, that is the peak value of the input audio signal. The desired decision is to find the optimal state sequence $\tau_{1:T}^*$ which refers to the observational data and formulated in eq. (9) [10].

$$g(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n,k)|$$　　　　　　　　(9)

Therefore, onset detection process does not require frequency information, then all output STFT magnitude at each frequency bin is summed to obtain the total magnitude of each time frame. If the audio signal under study is a complex audio signal, which many instruments played at once, then the sum of all magnitudes are only performed in the frequency range of instruments to analyze the beat.

The probability of the observed to occur if state-t^{th}-happens to $P(o_t|\tau_t)$ is divided into two probability values, ie the probability of the observed onset frame t is a $P(o_t|\tau_t=1)$ and the probability of the observed if frame to- t-th is not an onset frame P (ot | πt ≠ 1), and second overall probability value amounted to 1.

The probability of the observed data is an onset frame $P(o_t|\tau_t=1)$ is determined from the results of elevation normalized output extraction of STFT process' magnitude. Higher the peak value of the magnitude of a frame t^{th} are observed, the more likely the frame is a frame onset. Contrarily, if the value of magnitude lower, it is likely that the tth observed frame is not an onset frame

.

$$P(o_t|\tau_t = 1) = g(n) \tag{10}$$

$$P(o_t|\tau_t \neq 1) = 1 - g(n) \tag{11}$$

In eq. (10) and (11), it can be said that the calculation of the probability of the observations can be considered as a fixed thresholding with a value of 0.5 is often used in the conservative onset detection methods. If the peak value is greater than 0.5, the peak is an onset, and vice versa.

The calculation of the state transition probability is denoted by the symbol Ps,u which is the probability of s state is changed to state u or $P(\tau_t=s| \tau_{t+1}=u)$, where $s,u \in \{1,2,…,S\}$. Because of hidden state variables represent the frame index calculated from the onset of the previous frame sequence, then it is likely that the only possible state changes from s to $s+1$ or 1. Ps,u may happen only $Ps,s+1$ dan $Ps,1$ which means that the next frame is a frame onset. Illustration of the transition state of the HMM method is illustrated in Figure 6.
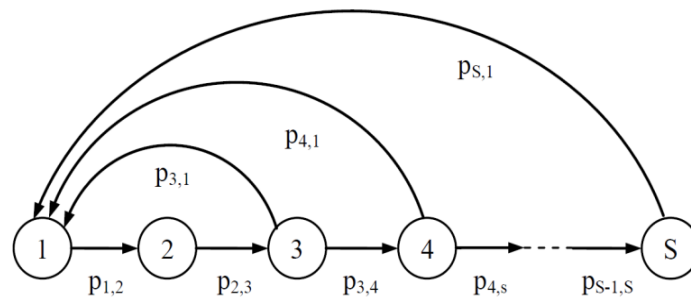

Figure 6. Illustration Relationship between State Transition Opportunities

$Ps,1$ values, is modeled as Gaussian probability distribution that has a peak at an average frame distance between two successive onset. The average value can be obtained from the value of music tempo on the commonly used–composition of Gamelan, eg 60 bpm tempos means that in 1 minute on average there are 60 beats.

A simple example of modeling the distribution of the state transition can be seen in Fig. 7, with a number of state of 20 and an average of 10 Gaussian distribution, which means most likely that after the 10th frame to the state transitions to state 0 (a10, 0 = 0.99).
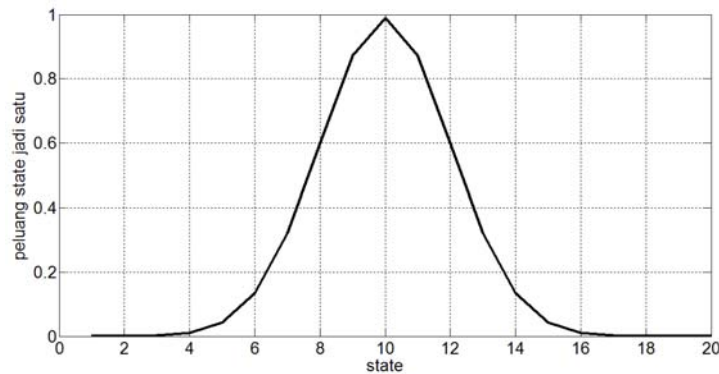

Figure 7. Simple Transition Distribution Model

While observational data o_t is defined as the peak value detection onset of the feature extraction process results when the tth τt state of the following probability distribution $P(o_t|\tau_t)$.. Illustration of the observed relationship with the hidden variable state can be seen in Figure 8.
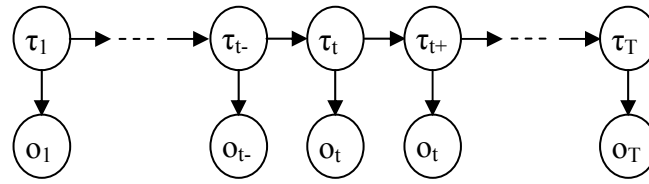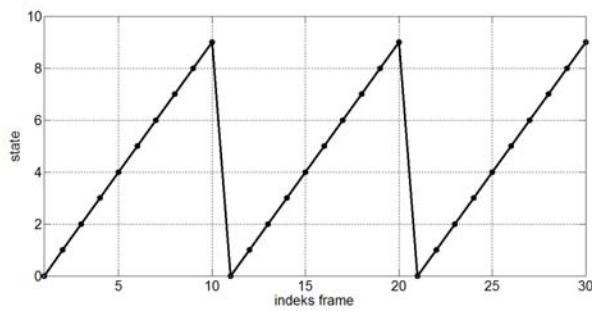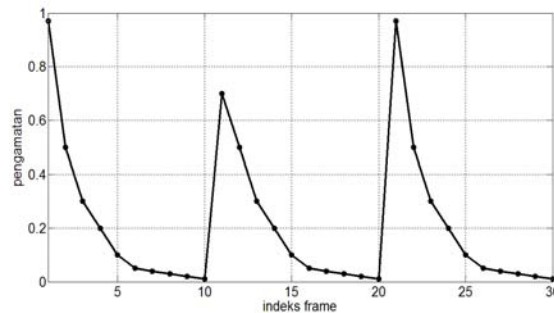
Figure 8. Graph Acyclic trending HMM

A simple example to illustrate again the corresponding relationship with the hidden state of observational data can be seen in Figure 9. When the value of observational data on a frame showing a peak or a high value, then the data should be hidden state corresponds to the state 0 in the frame.

(a)

(b)

Figure 9. Illustration of Data Observations and Hidden State; (a) Hidden State (b) Data Observations

Decisions that is going to be achieved from HMM beat detection method is to find the optimal state sequence $\tau_{1:T}^*$ which refers to data observation o_t. The integration process and the transition probability value of observational data is illustrated in Figure 10. While simple example of merging the transition distribution model Figure 7 with quite a variety of observational data can be seen in Figure 10.
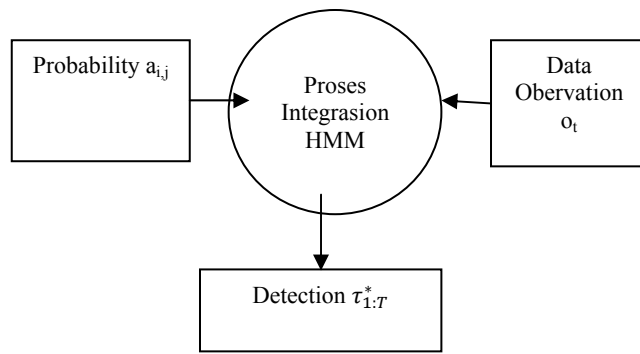
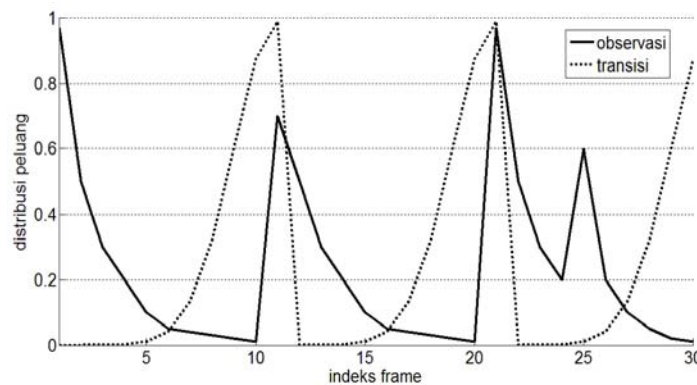Figure 10. Integration Simple HMM for Onset Detection

Figure 11. Distribution of Observations Integration and Transition

In Figure 11, after the onset frame is detected, then the state of transition distribution model move again from state 0, and so on until the last frame. Figure 11, it can also be seen that the existence of a false peak in the 25th frame. However, because the frame is the value of the state transition probabilities into state 0 is low,   therefore the counting of the frame state continues.

If there are many variations of tempo, all values of tempo variations in the calculations can be included on Ps, 1 as written in eq. (11).

$$P_{s,1} = \frac{1}{K} \sum_{k=1}^{K} \exp\left(-\frac{(s-\mu_k)^2}{2\sigma_k^2}\right) / \sqrt{2\pi\sigma_k^2} \qquad (12)$$

where K is the number of possible variations in tempo and μ_k is the value of music to variations in the *kth* period-. σ_k is the standard deviation of the value of music to the *kth* period.

Illustration value of $P_{s,1}$ with two variations of tempo can be seen in Figure 12 which shows the value of the probability of a subsequent frame if the previous onset frame have an s-state . Two tempo variations expected to occur are 60 bpm and 120 bpm, which means that the average onset frame appears at a distance of every 100 frames and 50 frames, given the distance between each frame is as wide as 10 ms hop size.
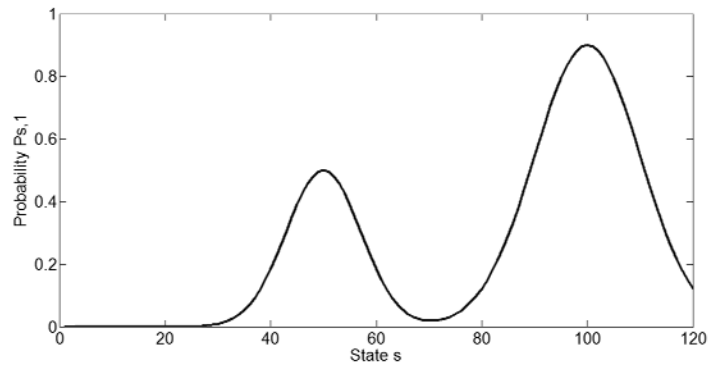
Figure 12. Illustration Probability State Being Onset

The calculation of the latter is the initial value of the assumed probability is the probability with uniform distribution among all N number of states that may arise.

$$P(\tau_1 = n) = \frac{1}{N} \; untuk \; n = 0, ..., N - 1 \tag{13}$$

By locating the frames that have a state value of 1 from the eq. (13), the performance of HMM method can be measured and compared to the performance of the method of Spectral Flux.

In this study, the input value of the musical period is obtained from the calculation peak distance of first 4 seconds, so when the orchestra is including the slow tempo (tempo less than 60 bpm), the first 4 seconds of the calculation can be obtained at least 2 peaks. While the number of state are included also has the same value, ie N = 400 which is equivalent to 4 seconds (hop size is used as the distance between frames is 10 ms).
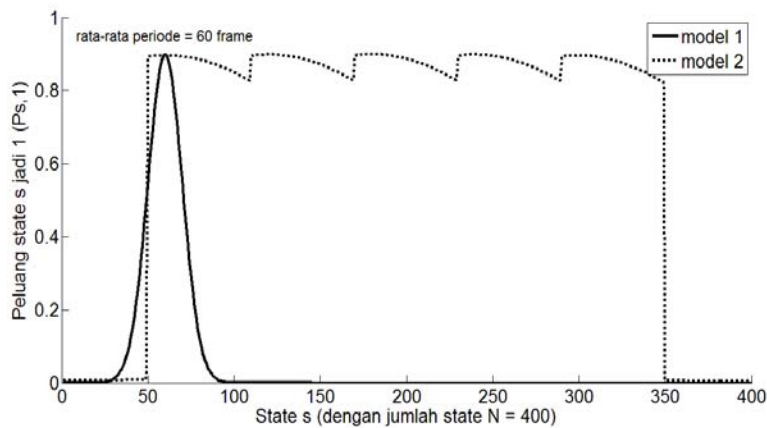

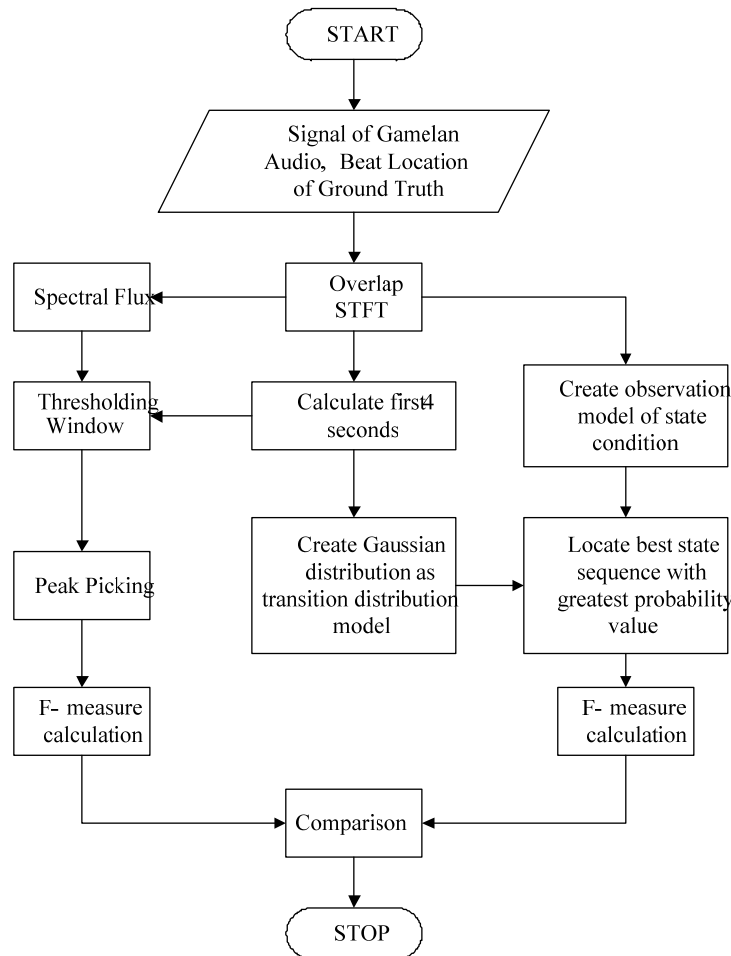Figure 13. Illustration of Two Transition Distribution Model

Figure 14. Overall Flow Chart Method

The process of modeling distribution of the transition state of this research is made of two possible models. The first model is a Gaussian distribution with a peak average value obtained from the average distance between the peaks in the first 4 seconds. While the second model, many peaks Gaussian distribution are made, on which the peaks are multiples of the average distance between the peaks in the first 4 seconds. Illustration of the two models can be seen in Figure 13. Selection of models made at the end of the frame that has the greatest probability value that fit with observed data from all frames on the total multiplication eq. (14). While the overall flow diagram of the complete method is shown by Figure 14.

$$\tau_{1:T}^{*} = \underset{\tau_{1:T}}{\operatorname{argmax}} \left[P(o_1|\tau_1)P(\tau_1) \prod_{t=2}^{T} P(o_t|\tau_t)P(\tau_t|\tau_{t-1}) \right] \qquad (14)$$

with T is the total number of frames of the observed audio signal is, $P(\tau_1)$ is the value of initial probability, $P(\tau_t|\tau_{t-1})$ state transition probabilities, and $P(o_t|\tau_t)$ is the probability of the observed values with state requirements of tth. Directed acyclic graph showing the relationship between the observational data with state frame from the first frame to the last frame. The calculation of the value of the observations used in the HMM method in this study is the extraction of features from the output magnitude STFT.

3.   Results and Discussion
3.1. Performance Measurement

To test the performance of the onset detection system yields F-measure is calculated parameter which is a major requirement in the provision MIREX (Music Information Retrieval Evaluation Exchange) [8].

$$P = Precision = \frac{n_{tp}}{n_{tp}+n_{fp}} \quad (15)$$

$$R = Recall = \frac{n_{tp}}{n_{tp}+n_{fn}} \quad (16)$$

$$F - measure = \frac{2PR}{P+R} \quad (17)$$

with n_{tp} is the number of true positives (number of beats right), n_{fp} is the number of false positives (wrong number of beats) and n_{fn} is the number of false negatives (number of beats that are not detected).

As for the location of a beat would be true if the actual beats still within tolerance of ± 70 ms of the detection results. This is in accordance with the specifications of MIREX and to make room for manual labeling process which may be less accurate. If there is more than one beat detection results within the tolerance limits, then only one is counted as a true beat detection and others counted as false positives. If a beat detection is right on the boundary between the two beat the real location, it is considered there is a true beat detection and a false negative.

In this experiment, the method of HMM (Hidden Markov Model) and the method of Spectral Flux compared performance on synthetic track data and acoustic instruments played by balungan group, namely demung, Saron, and Peking. The song is played has the same notation and the distance between the beats are almost the same anyway.

We generated two types of gamelan sound for testing:
1. Synthetic. Each gamelan note was recorded and the ensemble was played using computer with gamelan note direction.
2. Acoustic. Gamelan ensemble was played by the players and was recorded.

Because the data track is tested synthetic track, then all the data this song has a fixed tempo, so the F-measure value measured is high. Variation of the F-measure on experiments with synthetic tracks influenced by the presence of a single recording signal amplitude variations.

Tabel 1. Performance comparison on synthetic instrument data

| Type	Name of song	Instrument	Time (s)	F measure Spectral Flux	HMM
Sintetic	Manyar Sewu	Saron	30	91%	98%
Sintetic	Manyar Sewu	Demung	54	72%	98%
Sintetic	Manyar Sewu	Saron & Demung	73	85%	99%
Acoustic	Manyar Sewu	Saron	29	84%	88%
Acoustic	Manyar Sewu	Demung	52	66%	72%
Acoustic	Manyar Sewu	Saron & Demung	70	82%	96%
Acoustic	Ayak Talu	Saron	82	91%	95%
Acoustic	Manyar Sewu	Orkestra	75	55%	73%
Acoustic	Gugur Gunung	Orkestra	47	36%	62%
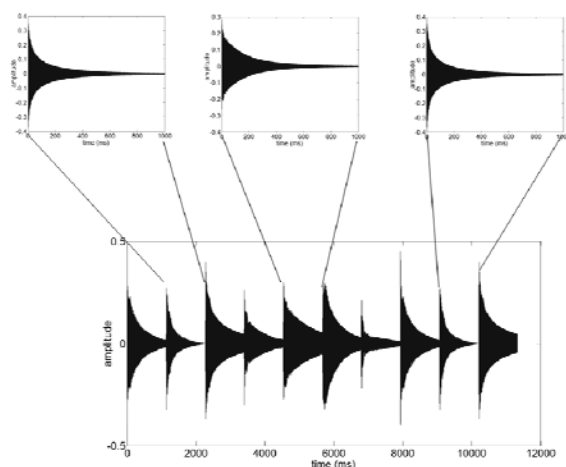
Figure 15. Illustration song Synthetic Process Data
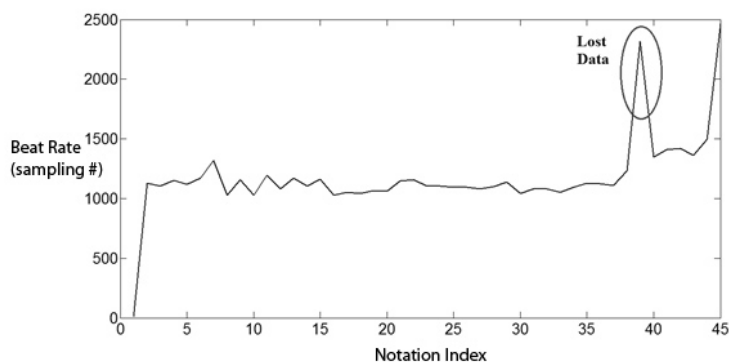
Figure 16. Tempo Music at Data song demung

Significant improvement occurred in the data demung synthetic track, because there are lot of false negatives, as shown in Figure 15.

On test, the performance gained by 98% due to the irregularity demung beats which can be shown in Figure 16. On these data, there is a shortage of beats after the 38th notation which make onset detector to be likely to go wrong. This was evidenced by the F-measure obtained by Spectral Flux method only 91%, far below the performance of HMM methods.

In this experiment, the Manyar Sewu song data is used. It is played in the orchestra and recorded immediately. Because there are many instruments being played, then the assessment of system beats (beat tracking) the frequency of the screening process should be done in advance and –beat detection focused on signals from the instrument balungan Saron or demung, because the song Manyar Sewu, Saron instruments and demung played with a single beat corresponding notation song.

The diversity of instruments orchestra played the Manyar Sewu track, generates very various music signals of very varied. STFT Magnitude results have the height value diversity so is not an uncommon onset detection methods perform error detection.

4. Conclusion

In order to construct a robust instrument extraction from music ensemble, From all the experimental results in this study for the assessment of the performance of beat detection signal onset balungan on Gamelan music, detection of the onset using HMM method has a high

performance up to 89% on a played single instrument song data. While the song data with many instruments, achieving 91% accuracy. Single instrument the song data which often change tempo, HMM detector has a performance of up to 95%. HMM method can improve the performance of the F- measure is better 10% when compared to the use of Spectral Flux.

References
[1]　YK Suprapto, M Hariadi, MH Purnomo. Traditional music sound extraction based on spectral density model using adaptive cross-correlation for automatic transcription. *IAENG International Journal of Computer Science.* 2011; 38(2).
[2]　DP Wulandari, YK Suprapto, MH Purnomo. *Gamelan music onset detection using Elman network.* IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Tianjin, China. 2012: 91-96.
[3]　Sumarsam. *Cultural Interaction and Musical Development in Central Java.* The University of Chicago Press. ISBN 0-226-78011-2. 1992-1995.
[4]　DP Wulandari, A Tjahyanto, YK Suprapto. *Recurrent neural networks for Gamelan music onset detection.* Proc. of the 3rd Int. Conf. on Applied Technology, Science, and Arts, Surabaya. 2011: 481-486.
[5]　M Caetano, JJ Burred, X Rodet. *Automatic segmentation of the temporal evolution of isolated acoustic musical instrument sounds using spectro-temporal cues.* Proc. of the 13th Int. Conf. on Digital Audio Effects (DAFx-10), Graz, Austria. 2010.
[6]　A Holzapfel, Y Stylianou, AC Gedik, B Bozkurt. Three dimensions of pitched instrument onset detection. *IEEE Transactions on Audio, Speech, and Language Processing.* 2012; 18(6): 1517-1527.
[7]　S Dixon. *Onset detection revisited.* Proc. of the 6th Int. Conf. on Digital Audio Effects (DAFx-06), Montreal, Canada. 2006.
[8]　JP Bello, L Daudet, S Abdallah, C Duxbury, M Davies, MB Sandler. A tutorial on onset detection in music signals. *IEEE Trans. on Speech and Audio Processing.* 2005; 13(5): 1035-1047.
[9]　H Thornburg. Detection and Modelling of Transient Audio Signals with Prior Information. *Ph.D. Dissertation.* Center for Res. In Music and Acoust., Stanford University, California. 2005: 51-121.
[10]　S Bock, M Schedl. *Enhanced beat tracking with context-aware neural networks.* Proc. of the 14th Int. Conf. on Digital Audio Effects (DAFx-11), Paris, France. 2011.
[11]　A Graves. Supervised Sequence Labelling with Recurrent Neural Networks. *Ph.D. Thesis.* Technische Universitat Munchen, Germany. 2008: 13-39.
[12]　N Degara, MEP Davies, A Pena, MD Plumbley. Onset event decoding exploiting the rhythmic structure of polyphonic music. *IEEE Journal of Selected Topics in Signal Processing.* 2011; 5(6): 1228-1239.
[13]　M Muller, DPW Ellis, A Klapuri, G Richard. Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing.* 2011; 5(6): 1088-1110.